

# Foreword: 1st International Workshop on High Performance Computing for Big Data

Kamer Kaya  
Computer Science and Engineering  
Faculty of Engineering and Natural Sciences  
Sabancı University  
Istanbul, Turkey  
Email: kaya@sabanciuniv.edu

Buğra Gedik  
Department of Computer Engineering  
Bilkent University  
Ankara, Turkey  
Email: bgedik@cs.bilkent.edu.tr

Ümit V. Çatalyürek  
Department of Biomedical Informatics  
The Ohio State University  
Ohio, USA  
Email: catalyurek.1@osu.edu

**Abstract**—The 1st International Workshop on High Performance Computing for Big Data (HPC4BD) is held on September 10, 2014 in concordance with 43rd International Conference on Parallel Processing (ICPP-2014). The workshop aimed to bring high performance computing (HPC) experts and experts from various application domains together to discuss their Big Data problems. There were four works accepted to be presented in this year’s workshop. This foreword presents a summary of the them.

## I. INTRODUCTION

Processing large datasets for extracting information and knowledge has always been a fundamental problem. Today this problem is further exacerbated, as the data a researcher or a company needs to cope with can be immense in terms of volume, distributed in terms of location, and unstructured in terms of format. Recent advances in computer hardware and storage technologies have allowed us to gather, store, and analyze such large-scale data. However, without scalable and cost effective algorithms that utilize the resources in an efficient way, neither the resources nor the data itself can serve to science and society at its full potential.

Analyzing Big Data requires a vast amount of storage and computing resources. We need to untangle the big, puzzling information we have and while doing this, we need to be fast and robust: the information we need may be crucial for a life-or-death situation. We need to be accurate: a single misleading information extracted from the data can cause an avalanche effect. Each problem has its own characteristic and priorities. Hence, the best algorithm and architecture combination is different for different applications.

The workshop aimed to bring people who work on data-intensive and HPC in industry, research labs, and academia together to share their problems posed by the Big Data in various application domains and knowledge required to solve them.

## II. HPC4BD’2014 PRESENTATIONS

- James Arnold Faeldon, Karen Espana, and Delfin Jay Sabido, *Data-Centric HPC for Numerical Weather Forecasting*: The authors demonstrate the integration of data analytics techniques for weather forecasting. They focus to manage and process large amounts of data arriving in time-sensitive streams from remote sensors and weather models and to reduce data analysis cycles significantly so that in

practice, the domain experts can make timely improvements on the model.

- Omer Baluch and Todd Eavis, *Soft Real-time OLAP: Exploiting Modern Hardware without Breaking the Bank*: The authors present a framework that reduces data integration costs and provides soft real-time OLAP functionality. They utilize hot data partitions that absorb and isolate incoming tuple streams so that integration into existing partitions or indexes is not required. They also apply a variety of multi-core processing techniques in order to significantly accelerate view construction and query resolution.
- Yingbo Cui, Xiangke Liao, Xiaoqian Zhu, Shaoliang Peng, and Bingqiang Wang, *Parallel Sequence Alignment for High-throughput Sequencing Data on Tianhe-2 Supercomputer*: The authors target an important problem and describe a BWA-based parallel sequence alignment tool to analyze of high-throughput sequence data on Tianhe-2. They propose strategies to reduce I/O overheads, utilize the processing capabilities of Intel Xeon Phi’s and distributed computing power of the supercomputer.
- Dinesh Kumar, Arun Raj, Deepankar Patra, and Dharanipragada Janakiram, *GraphIVE: Heterogeneity-Aware Adaptive Graph Partitioning in GraphLab*: A balanced work distribution on parallel computing can be hard to achieve, especially for irregular data such as graphs, since the graphs with small-world power-law characteristics pose additional challenges. The authors focus on this problem and utilize GraphLab to obtain significant improvement in performance with a novel dynamic work distribution technique.

## III. ACKNOWLEDGEMENTS

We thank all the authors for submitting their high-quality work to HPC4BD and their workshop presentations.

We also thank to the program committee members, Berkant Barla Cambazoglu, Mahantesh Halappanavar, Nilesch Jain, Heng Ji, Vana Kalogeraki, Tevfik Koşar, Tahsin Kurç, Kamesh Madduri, Ioan Raicu, Siva Rajamanickam, Sanjay Ranka, Erik Saule, Scott Schneider, Bora Uçar, and Peter R. Pietzuch. Their reviews helped a lot during the paper selection process.

Last but not least, we thank to ICPP Workshop Chairs Pavan Balaji and Anne Benoit for their help and making the organization easier for us.

# HPC4BD 2014 Workshop

## Program Committee

Berkant Barla Cambazoğlu, Yahoo Research  
Mahantesh Halappanavar, Pacific Northwest National Laboratory  
Nilesh Jain, Intel Labs  
Heng Ji, Rensselaer Polytechnic Institute  
Vana Kalogeraki, Athens Uni. of Economics and Business  
Tevfik Koşar, University of Buffalo  
Tahsin Kurç, Stony Brook University  
Kamesh Madduri, Pennsylvania State University  
Ioan Raicu, Illinois Institute of Technology  
Siva Rajamanickam, Sandia National Laboratories  
Sanjay Ranka, University of Florida  
Erik Saule, University of North Carolina Charlotte  
Scott Schneider, IBM Research  
Bora Uçar, CNRS and LIP, ENS Lyon  
Peter R. Pietzuch, Imperial College London